Conference or Workshop Item:

Laird, I., Murphy, Damian Thomas orcid.org/0000-0002-6676-9459 and Chapman, Paul (2012) Energy-based calibration of virtual performance systems. In: 15th International Conference on Digital Audio Effects (DAFx12), 17-21 Sep 2012.

ENERGY-BASED CALIBRATION OF VIRTUAL PERFORMANCE SYSTEMS

Iain Laird, Dr Paul Chapman
Digital Design Studio
Glasgow School of Art
Glasgow, UK
I.Laird1@gsa.ac.uk
p.chapman@gsa.ac.uk

Dr Damian Murphy
Audio Lab, Department of Electronics
University of York
York, UK
dtm3@ohm.york.ac.uk

ABSTRACT

A Virtual Performance System (VPS) is a real-time 3D auralisation system which allows a musician to play in simulated acoustic environments. Such systems have been used to investigate the effect of stage acoustics on the performance technique of musicians. This article describes the process of calibrating a VPS using energy-based quantities and goes on to verify this technique by comparing known acoustic quantities measured in a test space with a virtual version of the same space. This work has demonstrated that calibrating a VPS using metrics based on Support will result in an accurate simulation of a test space according to known acoustic metrics such as T30. A comparison of quantities referring to earlier parts of the response, such as Early Decay Time (EDT), show some errors which are thought to be caused by the non-anechoic nature of the reproduction space.

1. INTRODUCTION

The acoustic design of a stage is a critical part of constructing a successful performance space. It has been demonstrated that musicians will often adapt their playing technique (sometimes unconsciously) in reaction to the acoustic response they hear on stage [1]. When studying the effect of stage acoustics on performers, a primary concern is the venue in which the study takes place. Naturally, it is preferable to study musicians playing in actual concert hall environments; however hall hire costs and the lack of controllable variables make many experiments limited or impractical. It is therefore necessary to consider a laboratory set up which can emulate the 3D soundfield experienced by a performer playing on stage in response to the sound of their instrument in real-time.

Gade [2], Ueno [1] and Brereton [3] have previously conducted studies in virtual versions of concert halls which utilise a 3D auralisation system to recreate the sensation of playing on stage for a number of test subjects. While Gade's virtual acoustic environment was created using a combination of delay lines and loudspeaker-based reverberation, recent systems [1, 3] rely on convolution engines to apply a measured or modelled room response to the direct sound of the musical instrument which is then played back to the musician in real-time.

Systems such as this however remove the musician from their natural performing environment and place them in what is often reported to be an unnatural sounding approximation of stage acoustics. It is therefore crucial that the VPS can be verified to ensure the simulated acoustic environment is a fair representation of the target space. This article describes the process of verification using well known energy-based quantities. It will begin with a brief review of the important elements of stage acoustics and the methods used to recreate virtual stage acoustic conditions using a VPS. It will then describe an experiment which aims to reproduce the acoustic conditions of a test space and objectively verify its correct operation with reference to critical elements of concern for soloist musicians.

1.1. Soloist concerns

Gade's [2] initial studies began by interviewing performers regarding their main areas of concern with regards to stage acoustics. He uncovered a pattern in the use of vocabulary and defined a set of primary acoustic concerns for soloists highlighted below.

<u>Reverberance</u> – Is mainly perceived during breaks or shift of tone played, since it sustains the tones just played. It binds adjacent notes together, can blur details in the performance and may give a sense of response from the hall.

<u>Support</u> - Makes the musician feel that they can hear themselves and that it is not necessary to force the instrument to develop the tone. It can be felt during the onset of tones and is therefore believed to be related to properties different from reverberance.

<u>Timbre</u> - The influence of the room on the tone colour of the instrument and on the balance in level in different registers.

<u>Dynamics</u> – Describes the dynamic range obtainable in the room and the degree to which the room obeys the dynamic intentions of the player.

He found that audible levels of early reflections are preferred by the majority of soloists and that this subjective effect could be described as Support. Gade went on to demonstrate how Support could be quantified in terms of energy.

1.2. Stage acoustic parameters

Support (ST) is an objective quantity which determines the energy in an impulse response (in dB) within certain time intervals with reference to the direct sound. It is often used to determine how the energy within these time regions assists a performer's own efforts. Two main ST quantities exist which each refer to different time regions of an impulse response and are thus linked to assessing different perceptual effects.

$$ST_{early} = 10log_{10}\left(\frac{\int_{20ms}^{100ms} h^2(t)dt}{\int_{0ms}^{10ms} h^2(t)dt}\right) \quad (1)$$

$$ST_{late} = 10log_{10}\left(\frac{\int_{100ms}^{1000ms} h^2(t)dt}{\int_{0ms}^{10ms} h^2(t)dt}\right) \quad (2)$$

Where h^2 denotes the squared sound pressure of the impulse response measured on stage.

ST_{early} (1) is generally used to describe the degree of mutual hearing in ensembles while ST_{late} (2) is indicative of the degree to which the reverberant sound supports the musician's own efforts.

A standard measurement methodology is defined in ISO 3382 [4] however there remains ongoing discussion as to the reliability of ST parameters with many authors suggesting various alternatives for integration limits and source-receiver positions [5].

Reverberance can be quantified using standard reverberation time quantities such as RT60, T30 and EDT. Reverberation time (RT60) is defined as the time it takes for the measured sound pressure level in a room to attenuate by 60dB after a steady-state sound source has been switched off. T30 defines RT60 by extrapolation (using linear regression) over a 30dB range between -5dB and -35dB. EDT similarly defines RT60, now based on the first 10dB of attenuation and has been found to be linked to the perception of reverberance in a space [4].

2. SOUNDLAB AND VPS

A VPS system was constructed in the Arup/DDS SoundLab situated in Glasgow, UK. The SoundLab is a dedicated auralisation suite used for research and commercial purposes. The space is 4m (l) x 6m (w) x 2.5m (h), is heavily acoustically treated and is floated on springs in order to acoustically isolate the space. The SoundLab has a measured L_{AF90} (level exceeded for 90% of measurement time (5 mins)) of approximately 20dBA with all equipment in the SoundLab switched on. The average T20 of the space at 500Hz is approximately 0.15s. The SoundLab features a 12-channel periphonic loudspeaker array comprising of Yamaha MSP5A loudspeakers in addition to a Tannoy TS12 subwoofer. The speaker system has been equalised to ensure an optimum frequency response at the sweetspot and an equal level contribution measured at the sweetspot.

The VPS uses a microphone placed near the musician's instrument to pick up the direct sound of the instrument. The sound is convolved with an ambisonic [6] impulse response which can be measured using an ambisonic microphone or modelled using acoustic modelling software. The convolution is achieved using two Reaverb convolution plug-ins hosted in the Reaper Digital Audio Workstation (DAW) [7]. The resultant auralised material is then decoded using a Decopro ambisonic decoder [8] and played back in real-time to the musician sat in the centre of the speaker array.

Previous research [9] has found that it is not practical for the VPS to reproduce the direct sound or 1st reflection from the floor, as these elements are present naturally when a musician is playing their instrument in any space. Furthermore, any processing latency restricts the system's ability to accurately reproduce reflections within a certain time frame. Therefore these elements are discarded from the impulse responses used in auralisation. The resulting redundant propagation time is truncated by the appropriate amount in order to reduce overall system latency (described in Section 3.3).

It was also demonstrated in [9] that the level of acoustic response should vary in accordance with the level of the soloist's instrument and should not auralise quiet sounds to the same degree as louder sounds. Errors in the level of simulated acoustic response can contribute to a noticeable „PA Effect' whereby the musician feels as if they are playing through a PA system in a space rather than acoustically.

3. EXPERIMENT

An experiment was devised which aimed to test the operating performance of a VPS in reference to reverberance and dynamics. 1st order Ambisonic impulse responses were captured in a test space (described in Section 3.2) from a nearby loudspeaker using well known Swept Sine Wave techniques [10]. This data was used to create a virtual version of the space which was then calibrated using an energy-based quantity. After calibration, an identical survey was undertaken in the VPS. Quantities related to reverberance (EDT and T30) and dynamics (ST_{early} and ST_{late} over a range of SPL levels) were extracted and compared with data obtained in the test space to assess the quality of the simulation. ST_{early} and ST_{late} are used in this experiment as calibration parameters rather than to assess musician perception of acoustic support. It provides a convenient quantification of the energy in particular time regions of the acoustic response in reference to the direct sound which can be easily recreated in the VPS. A number of deviations from the standard measurement methodology were made, mainly relating to source directivity and source-receiver positions. For instance, ISO 3382 requires the use of an omnidirectional loudspeaker and microphone positioned 1m apart where this experiment required the use of a directional loudspeaker in close proximity to a Soundfield microphone.Thus in order to distinguish these quantities from ST_{early} and ST_{late} they are renamed E_{early} and E_{late} respectively.

E_{early} and E_{late} were also used to quantify the energy in the defined time windows over a range of sound source levels. This gave an indication that the VPS was providing the correct level of acoustic response over a wide source dynamic range.

3.1. Sound Source levels

In order to assess E_{early} and E_{late} over a known dynamic range, it was necessary to measure the Sound Pressure Level (SPL) generated by the loudspeaker.

A logarithmic sine sweep was generated in Matlab and output as a 32-bit floating-point, 44.1kHz wave file with a duration of 10s, swept from 1Hz - 22050Hz. The inverse sweep was also generated to enable extraction of the impulse responses from the recorded signal via convolution [10].

A Genelec 1029A loudspeaker was placed on a tripod in the SoundLab at a height of 85cm above the floor. A B & K 2231 Sound Level Meter (SLM) was positioned 10cm in front of the acoustic centre of the speaker. The loudspeaker was connected to a soundcard (Behringer ADA8000 connected to an M-Audio Profire Lightbridge) and Macbook running the Reaper DAW. Reaper was used to play back sine sweeps at particular gain settings which were attenuated in 6dB steps per sweep. The SLM was used to record the $L_{A,F,MAX}$ (Maximum A-weighted SPL, Fast time weighting) of the sine sweeps at each gain setting. The maximum level was obtained from a small number of repetitions of the sine sweep and was used in favour of other time-dependent acoustic metrics. The range of sound source SPLs was between 106.0dBA and 52.7dBA

3.2. Impulse response capture in test room

The equipment set up is shown in Figure 1. The loudspeaker was positioned in a large test room at a height of 85cm from the floor. A Soundfield ST350 microphone was placed 20cm directly behind the loudspeaker at a height of 128cm from the floor. This source-receiver configuration was set to approximate a geometric simplification of a seated wind musician. The loudspeaker was connected to the soundcard and laptop which played sine sweeps at the same gain settings as shown in Table 1. The Soundfield microphone was also connected to the soundcard which recorded the response of the room to this signal.

The test room was a large empty office unit on the 3rd floor of the Hub building at Pacific Quay in Glasgow. The space is rectangular in construction and measures approximately 53m (l) x 14m (w) x 4m (h). The space has a stainless steel floor and a large number of windows on each wall; the ceiling is also of steel construction with exposed building services over the full area. Small columns are located in the centre of the space along its length. The measurement equipment was set up approximately 8m from the North wall and 4m from the nearest column. The background noise in the space was measured with the SLM over a period of 15 minutes and was found to give an L_{AF90} of 39dB.

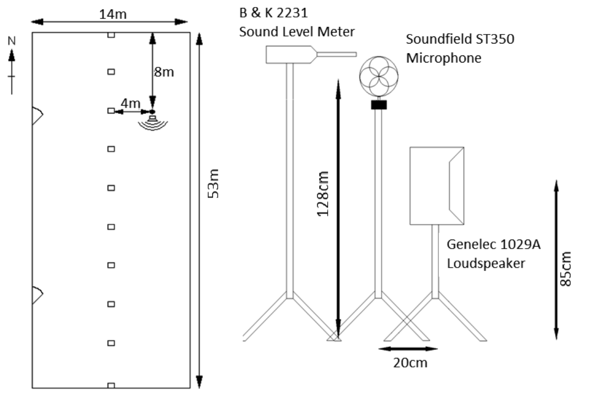

Figure 1: *Plan of test room on 3rd Floor of the Hub Pacific Quay, Glasgow. Source and receiver position and orientation are also shown.*

3.3. Construction and verification of VPS

The impulse responses were edited to remove the direct sound and floor reflection and then inserted into two Reaverb convolvers hosted in Reaper. The system soundcard was set to the lowest stable I/O buffer setting which was found to be 256 samples resulting in a processing latency of approximately 12ms.

The equipment set up described in 3.2 was reproduced in the SoundLab ensuring that the Soundfield microphone was positioned in the sweetspot of the ambisonic array. A Behringer ECM8000 measurement microphone was placed 10cm in front of the loudspeaker to feed the direct sound of the loudspeaker into the VPS. The impulse response with the highest signal to noise ratio obtained in the test room was edited as described and imported into two stereo Reaverb convolvers holding W, X, Y and Z channels of the ambisonic impulse response. Sine sweeps were played through the loudspeaker at the same level used to obtain that particular impulse response (106dB $L_{A,F,MAX}$). The auralised response to this signal was recorded at the sweetspot by the Soundfield microphone and the impulse responses derived as before.

By comparing the time of arrival of specific reflections in the impulse responses measured in the test room with those measured in the SoundLab it was possible to determine how much of the propagation time to discard in order for the reflections to occur at the correct time. Once aligned, the gain of the decoder output was adjusted to obtain the closest possible match in E_{early} and E_{late} measured in the test space. After the level of acoustic response had been calibrated, the sound source was adjusted to the other SPLs used in the test room. Impulse responses were measured for each source level and the E_{early} and E_{late} values derived as previously described.

4. RESULTS

Figure 2 compares E_{early} and E_{late} in the test room and virtual room. E_{late} in the measured and virtual case are well matched between source SPL of 106dB and 76dB SPL. With further attenuation of the source SPL it can be seen that the measured E_{late} starts to increase to an eventual gradient of approximately 6dB per attenuation step. The virtual version can be seen to follow this trend at 59dB. E_{early} in the virtual and measured cases react in a similar way with the increase beginning at much lower levels than in E_{late}. Between source SPLs of 106dB and 59dB the virtual E_{early} is typically 1dB higher than the measured E_{early}.

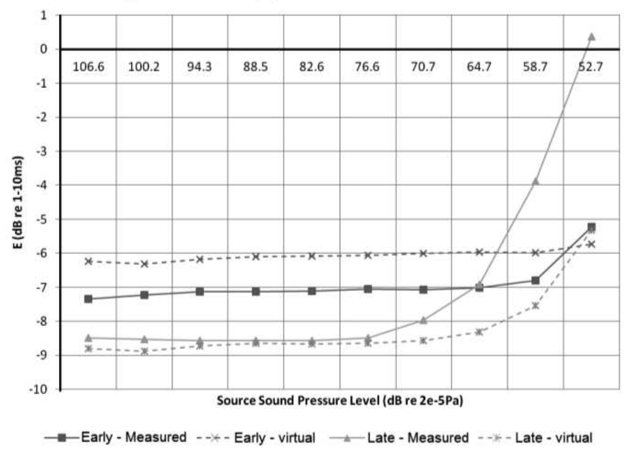

Figure 2*: plots showing measured (solid) versus virtual (dashed) E in early and mid time windows*

As the source level is gradually decreased, so too does the energy in each time window until it falls below the noise floor of the space. When this occurs the energy in this time window becomes constant, consequently further sound source level attenuation will cause an increase in E_{early}/E_{late}. In this experiment, the sound source level decrements by 6dB therefore once the energy in each time window falls below the noise floor it is expected E_{early}/E_{late} will increase by at a rate equal to this sound source level decrement. It can be observed that the E_{early}/E_{late} values start to increase at higher source SPL values in the measured space than in the virtual space. This is due to a lower background noise present in the virtual space.

Figure 3 compares EDT and T30 in measured and virtual rooms. 1/3rd octave results show that the virtual and test room T30 are highly correlated while EDT contains noticeable deviations particularly at 400Hz and 3.15kHz where the virtual space shows a reduced EDT. The maximum difference was observed to be approximately 2 seconds at 3.15kHz.

5. DISCUSSION

A visual comparison of measured and virtual impulse responses revealed additional reflections present in the first 20ms of the virtual response. This is thought to be caused by reflecting surfaces (TV screens and other equipment) in the SoundLab which may also account for the higher E_{early} values observed in the virtual response and the errors observed in EDT.

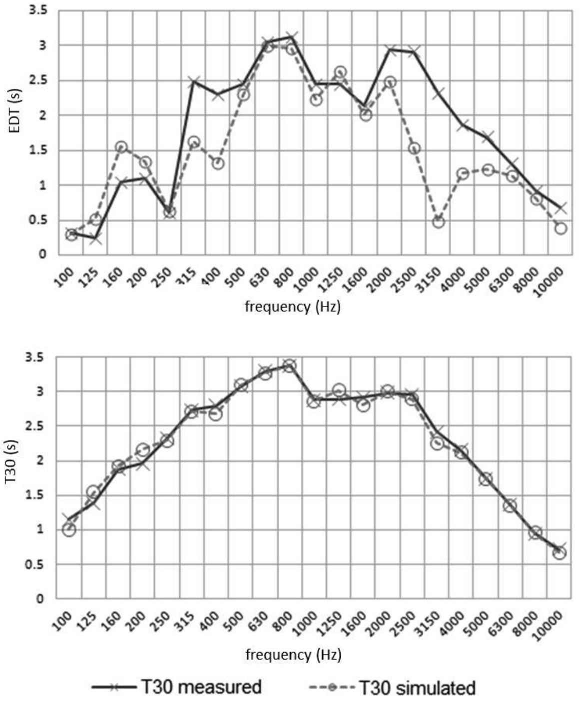

Figure 3*: plots comparing (a) EDT and (b) T30 of measured test room (solid) and the virtual version (dashed)*

Due to the non-anechoic nature of the SoundLab it is suggested that there is a lower limit for E_{early} simulations requiring reflections in this time window to be of higher amplitude than those present in the reproduction space. Studies involving fine adjustment of early reflections may need to be conducted in VPS systems housed in near anechoic conditions. Subjective testing will indicate how audible the observed differences are in the presence of masking noise created by a musician's instrument however this is expected to be highly dependent on the type of instrument studied.

6. CONCLUSIONS

The results of this experiment have indicated that energy based calibration of a VPS provides a reasonably accurate virtual response in terms of known acoustic quantities. The VPS was found to be less accurate when simulating the early part of the response due to interference from the SoundLab's own acoustic response.

An immediate concern is how well the VPS is simulating the direction of early reflections using a 1st order ambisonic system. A time-frequency directional analysis will be performed on the measured and simulated responses to indicate how well individual reflections are being spatialised. Furthermore, the results obtained in this study reflect a completely static musician however performing musicians will move and gesture while playing which will affect the achievable accuracy of the system. A system capable of tracking musician movements may improve accuracy.

7. ACKNOWLEDGEMENTS

This work was supported by the Digital Design Studio and Arup. Thanks to the Hub staff for allowing access to test spaces.

8. REFERENCES

[1] K. Ueno, K, Kosuke, K Lawai "Musicians' Adjustment of Performance to Room Acoustics Part 1: Experimental Performance and Interview in Simulated Soundfield." *19th International Congress on Acoustics*, Madrid, 2007

[2] A. C. Gade "Investigations of Musicians' Room Acoustic Conditions in Concert Halls. Part 1: Methods and Laboratory Experiments." *Acustica*, vol. 69, pp. 193-203, 1989

[3] J. Brereton, D. Murphy, D. Howard "A Loudspeaker-Based Room Acoustics Simulation for Real-Time Musical Performance." *Proc. Of the 25th UK AES Conference in assoc. with the 4th International Symposium on Ambisonics and Spherical Acoustics*, 2012

[4] BS EN ISO 3382-1:2009 "Acoustics – Measurement of room acoustic parameters", 2009

[5] A. C. Gade "Acoustics for Symphony Orchestras; status after three decades of experimental research" *Proceedings of the International Symposium on Room Acoustics*, Melbourne Australia, 2010

[6] P. Fellgett, "Ambisonics. Part One: General System Description", *Studio Sound, 1:20–22,40*, August 1975

[7] "Reaper DAW", Available at http://www.reaper.fm/, accessed Jan 2012

[8] "Gerzonic Ambisonic Decoders", Available at http://www.gerzonic.net/, accessed Jan 2012

[9] I. Laird, D. Murphy, P. Chapman, S. Jouan "Development of a Virtual Performance Studio with Application of Virtual Acoustic Recording Methods." *Proc. of the 130th AES Convention*, London, UK, 2012

[10] A. Farina "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique" *Proc. of the 108th AES Convention*. Paris, France, 2000